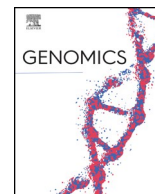




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Original Article

# SARS-CoV-2 transcriptome analysis and molecular cataloguing of immunodominant epitopes for multi-epitope based vaccine design

Sandeep Kumar Kushwaha<sup>a,\*</sup>, Veerbhan Kesarwani<sup>a,b</sup>, Samraggi Choudhury<sup>a</sup>, Sonu Gandhi<sup>a</sup>, Shailesh Sharma<sup>a</sup>

<sup>a</sup> DBT-National Institute of Animal Biotechnology (NIAB), Hyderabad 500032, Telangana, India

<sup>b</sup> Center of Bioinformatics, University of Allahabad, Prayagraj 211002, Uttar Pradesh, India



## ARTICLE INFO

## Keywords:

SAR-CoV-2  
 COVID-19  
 T- and B-cell  
 Peptide  
 Epitopes  
 Vaccine

## ABSTRACT

Genomics-led researches are engaged in tracing virus expression pattern, and induced immune responses in human to develop effective vaccine against COVID-19. In this study, targeted expression profiling and differential gene expression analysis of major histocompatibility complexes and innate immune system genes were performed through SARS-CoV-2 infected RNA-seq data of human cell line, and virus transcriptome was generated for T- and B-cell epitope prediction. Docking studies of epitopes with MHC and B-cell receptors were performed to identify potential T- and B-cell epitopes. Transcriptome analysis revealed the specific multiple allele expressions in cell line, genes for elicited immune response, and virus gene expression. Proposed T- and B-cell epitopes have high potential to elicit equivalent immune responses caused by SARS-CoV-2 infection which can be useful to provide links between elicited immune response and virus gene expression. This study will facilitate *in vitro* and *in vivo* vaccine related research studies in disease control.

## 1. Introduction

Corona viruses are a group of related viruses that are responsible for causing diseases ranging from the common cold to severe diseases like Middle East respiratory syndrome (MERS), SARS-CoV-2 in mammals and birds. SARS-CoV-2 is a positive sense single-stranded RNA virus belonging to the family *Coronaviridae* and subgenus *Sarbecovirus*. It is responsible for the widespread global pandemic causing an upper respiratory tract infection of humans [1]. SARS-CoV-2 virion ranges from approximately 50–200 nm in diameter [2]. SARS-CoV-2 is made up of four structural proteins known as the spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins. The nucleocapsid protein contains the viral RNA and the spike, membrane, envelope make up the viral envelope. The spike protein is responsible for the viral attachment with angiotensin-converting enzyme 2 (ACE2) receptors and facilitates entry into the host cells [3]. The ACE2 receptors are present in the goblet (secretory) cells of ciliated cells in the nose, back of the throat, lungs, gut, heart muscles and kidney which facilitates the hand to mouth transmission route. The viral RNA is released in the nasal cells when the transmembrane serine protease 2 (TMPRSS2) splits the spike proteins, and enters inside the cell, the viral genetic material replicates into millions. Seroconversion of SARS-CoV-2 took place within four days of

infection and found in most of patients by day 14. Moreover, persistence of specific IgG and antibody production was reported even after 2 years of infection [4]. Whereas limited serological details of SARS-CoV-2 are available at the moment, it is reported that a patient showed the presence of IgM after 9 days of infection, and later production of IgG after 2 weeks [5]. In an *in vitro* plaque testing with patient sera, it was confirmed that it is able to neutralize SARS-CoV highlighting the successful mounting of humoral response [6]. The current evidences have shown that Th1 immune response can be successful for controlling SARS-CoV [7] and may work for SARS-CoV-2 as well. The immune system responds in two ways, the innate immune response is a quick response which leads to inflammation, the immune cells kills the pathogens and induces adaptive immunity. The Adaptive immunity is slower and involves two types of cells the T-cell and the B-cell that react with the pathogens. T-cell kills the infected cells and the pathogens, while B-cells produce specific antibodies against the pathogens. Some of the T-cell and B-cell form the memory cells which are stored in the lymph nodes, in case the person is infected with the same infection memory cells produce antibodies against to combat the infection [8]. The epitopes are the antigenic regions of an antigens, causing an immune response which is identified by antibodies generated from T- and B-cells. T-cell epitopes present on the cell surface binds to the major

\* Corresponding author.

E-mail addresses: [sandeep@niab.org](mailto:sandeep@niab.org) (S.K. Kushwaha), [gandhi@niab.org.in](mailto:gandhi@niab.org.in) (S. Gandhi), [shailesh.sharma@niab.org.in](mailto:shailesh.sharma@niab.org.in) (S. Sharma).

<https://doi.org/10.1016/j.ygeno.2020.09.019>

Received 20 July 2020; Received in revised form 18 August 2020; Accepted 8 September 2020

Available online 10 September 2020

0888-7543/ © 2020 Published by Elsevier Inc.

histocompatibility complex (MHC) molecules. The MHC I molecules presents peptides of 8–11 amino acids in length, which are CD8+ T-cell epitopes. In contrast, MHC II molecules present longer peptides of 13–17 amino acids in length, which are CD4+ T-cell epitopes. The epitope-based vaccine development offers prospective advantages over the whole protein approach because the immune response against highly reserved epitopes over a widespread population can be used for the treatment of highly variable pathogens [9,10]. Various successful studies were reported for the epitope-based vaccine design against West Nile virus [11], dengue virus [12], chikungunya virus [13], shigellosis [14] etc. COVID-19 first cases were observed in Wuhan, China in December 2019, which seems to be the origin of SARS-CoV-2 virus. As of July 2020, there are more than 13.5 million confirmed cases, with 584 thousand deaths globally. Vaccines are mostly in the developmental stages, and no approved drug or vaccine triggering the immune response in the body against SARS-CoV-2 in the market.

In the present study, an integrated bioinformatics approach was used to identify expressed T- and B-cell epitopes from RNA-seq data of SARS-CoV-2 infection in human cell lines. To the best of our knowledge, no previous study has been reported a list of expressed T- and B-cell epitopes for multi-epitope based vaccine development. The specific objectives of this research study were: (1) Gene expression analysis of MHC gene and alleles, and innate immune system genes, (2) SARS-CoV-2 transcriptome construction, and annotation to explore expressed region of SARS-CoV-2 genome, (3) identification of potential T- and B-cell epitopes by using SARS-CoV-2 transcriptome, and (4) modelling and docking studies to explore structural compatibility of epitopes with MHC and B-cell receptors.

## 2. Materials and methods

### 2.1. SARS-CoV-2 data retrieval, processing and transcriptome assembly

Due to the recent outbreak of SARS-CoV-2, several countries have started to generate molecular resources to understand pandemic caused by SARS-CoV-2. We used publically available transcriptome data (PRJNA615032) of SARS-CoV-2 infection in normal human bronchial epithelial (NHBE) and human adeno carcinomic alveolar basal epithelial (A549) [15]. All the available data were download from the sequence read archive of NCBI database, and fastq-dump program of SRAToolkit [16] was used to extract fastq reads. Quality assessment and control of RNA-seq data was performed through the FastQC version 0.11.5 [17], MultiQC version 1.8 [18] and trimmomatic version 0.39 software [19]. All high-quality reads were mapped over SARS-CoV-2 isolate Wuhan-Hu-1(MN908947.3) by using HISAT2 version 2.1.0 on default parameters [20]. Samtools version 1.1.0 [21] and Bedtools version 2.26.0 [22] were used to extract all the mapped read from each sample and extracted reads were used to construct *de novo* transcriptome assemblies by using the Trinity assembler version 2.5.1 [23]. TransDecoder program [24] was used to generate protein sequence from assembled transcriptome. Kallisto, a pseudo aligner for bulk RNA-seq data alignment, was used for expression quantification [25]. Kallisto-sleuth pipeline was used for differential gene expression (DGE) analysis [26]. Human innate immune system genes were downloaded from innate database [27].

### 2.2. T-cell epitope prediction from SARS-CoV-2 transcriptome sequences

T-cell epitopes has potential to induce specific immune responses to control the disease which can be used as a key molecular resource for epitope-based vaccine design. NetCTL 1.2 program [28] was used to predict cytotoxic T-lymphocyte (CTL) epitopes from protein sequences, translated from assembled SARS-CoV-2 transcriptome. Epitopes were identified for all the 12 super types *i.e.* A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62 at the combined score threshold value 1.00. Antigenic potential of identified peptides were explored through

VaxiJen v2.0 [29] at the threshold value 0.4, and IEDB program (<http://tools.iedb.org/immunogenicity/>) was used to identify immunogenicity score for each identified epitopes. To predict MHC genes and alleles for all the identified epitopes, IEDB program mhci (<http://tools.iedb.org/mhci/download/>) and mhcii (<http://tools.iedb.org/mhcii/download/>) were used. Peptide length 9 and inhibitory concentration (IC50) value less than or equal to 200 nM were selected as parameters for the identification of MHC class I and II binding genes and alleles. IEDB program conservancy (<http://tools.iedb.org/conservancy/>) was used to identify conservation level of epitopes. IEDB population coverage tool (<http://tools.iedb.org/population/>) was used to analyse population coverage through predicted MHC alleles of epitopes [30], and peptide toxicity prediction was performed through the ToxinPred web server [31]. Cross-reactivity with host antigenic proteins might leads to adverse immune responses. Therefore, selected epitope sequences were checked for similarities with the human proteome sequences (*Homo sapiens*: GRCh38) through standalone NCBI BLAST similarity search tool.

### 2.3. B-cell epitopes prediction from SARS-CoV-2 transcriptome sequences

Sequence and structure based approaches were used to identify B-cell epitopes. In the sequence based approach, VaxiJen server was used to identify most antigenic proteins from translated transcriptome, and BepiPred-2.0 program [32] was used to identify B-cell epitopes from the identified antigenic protein. IEDB conformational B-cell prediction tool ElliPro (<http://tools.iedb.org/elliopro/>) were used to predict epitopes based on protein structure with the parameters PI (protrusion index) value 0.8 as minimum score, and 7 Å as maximum distance [33]. Protein sequences of SARS-CoV-2 transcriptome were showed strong sequence similarity with modelled 3D structure of SARS-CoV-2 genome at Zhang lab. Hence, we downloaded 24 structure of SARS-CoV-2 genome from Zhang lab (<https://zhanglab.ccmb.med.umich.edu/>) for structure-based epitope prediction. Epitopes identified by both the approaches were evaluated for toxicity, antigenicity, immunogenicity and cross-reactivity same as done for T-cell epitopes.

### 2.4. Molecular docking studies

Selected epitopes were studied for structural compatibility and interaction analysis with available MHC class I and II genes, and B-cell receptor structures. Protein structure of MHC class genes, and B-cell receptors were retrieved from RCSB Protein Data Bank (PDB) (<https://www.rcsb.org/>) in PDB format. Docking studies were performed using AutoDockTools, AutoDock Vina and CABS-dock server to identify structural compatibility between T- and B- cell epitopes and protein receptors [34–36]. Prediction of conformation compatibility between peptide and receptor in real world is even more complex and challenging. Therefore, initial screening of peptides for receptors were performed through Autodock, whereas CABSdocks was used for final binding and compatibility analysis. Protein complex visualization and hydrogen bonds were calculated through UCSF chimera [37] and LIG-PLOT software package [38], and also ensured the number of genuine hydrogen bonds through cavity prediction by using D3Packets server [39].

## 3. Results

### 3.1. Gene expression analysis of MHC genes and alleles

The MHC genes and alleles are the well explored regions of human genome because it is highly associated with different diseases, immune responses, and very well characterized at functional levels [40]. To explore expressed MHC gene and alleles, MHC gene expression profiling was performed by exploiting SARS-CoV-2 infected NHBE and A549 cell lines RNA-seq data. In our analysis, HLA-A variants (HLA-A\*01, HLA-

A\*02, HLA-A\*03, HLA-A\*11, HLA-A\*24, HLA-A\*25, HLA-A\*26, HLA-A\*30, HLA-A\*31, HLA-A\*32, HLA-A\*33, HLA-A\*34) were more expressed than HLA-B (HLA-B\*08, HLA-B\*15, HLA-B\*18, HLA-B\*37, HLA-B\*44, HLA-B\*45) and HLA-C variants (HLA-C\*03, HLA-C\*04, HLA-C\*06, HLA-C\*07, HLA-C\*12, HLA-C\*16). In MHC class II analysis, HLA-DMA\*01, HLA-DMB\*01, HLA-DOB\*01, HLA-DPB1\*08, HLA-DPB1\*75, HLA-DQA1\*01, HLA-DQB1\*06, HLA-DRA\*01 and HLA-DRB1\*13 gene variants were highly expressed to present exogenous antigens to CD4+ T cells. Cell line samples, expression of MHC class genes and alleles, and expression count (TPM) of alleles with respective to human cell line were given in Supplementary File 1, Tables S1 and S2). In the differential gene expression analysis of MHC genes by using mock and infected sample of NHBE cell lines, HLA-A\*24:02 is the only significantly expressed alleles between infected (TPM: 12331.79) and mock samples (TPM: 3822.033) (Supplementary File 1, Table S4).

### 3.2. Gene expression analysis of innate immune genes

Role of innate immune system is very important against respiratory diseases because human body is exposed to thousands of pathogens in daily life. Therefore, gene expression pattern information of innate immune gene can help us to improve our understanding about induced immune response for diseases. In differential gene expression analysis of human innate immune genes, 209 genes were differentially expressed between infected and mock samples of NHBE cell line. Differentially expressed gene, annotation, and involved pathways are given in Supplementary File 1, Tables S5 and S6. In differential gene expression analysis, myxovirus (influenza virus) resistance 1 (ENSG00000157601), ribosomal protein S19 (ENSG00000105372), phospholipid transfer protein (ENSG00000100979), signal transducer and activator of transcription 2 (ENSG00000170581), calmodulin 3 (ENSG00000160014), interferon (ENSG00000165949), SWI/SNF related gene (ENSG00000080503), intercellular adhesion molecule 1 (ENSG00000090339), TNFAIP3 interacting protein 1 (ENSG00000145901), adenosine deaminase (ENSG00000160710), and ubiquitin 1 (ENSG00000135018) are found most differentially up regulated in infected samples of NHBE cell line. In the list of down regulated genes, tumor necrosis factor (ligand) superfamily (ENSG00000121858), actin (ENSG00000075624), amyloid beta (A4) precursor protein (ENSG00000142192), activating transcription factor 3 (ENSG00000162772), clathrin (ENSG00000141367), phosphatidylinositol 3-kinase (ENSG00000078142), interleukin 1

(ENSG00000125538), a kinase (PRKA) anchor protein 10 (ENSG00000108599), zinc finger (ENSG0000015171) and apolipoprotein B mRNA editing enzyme (ENSG00000179750) were found in infected cells (Table 1).

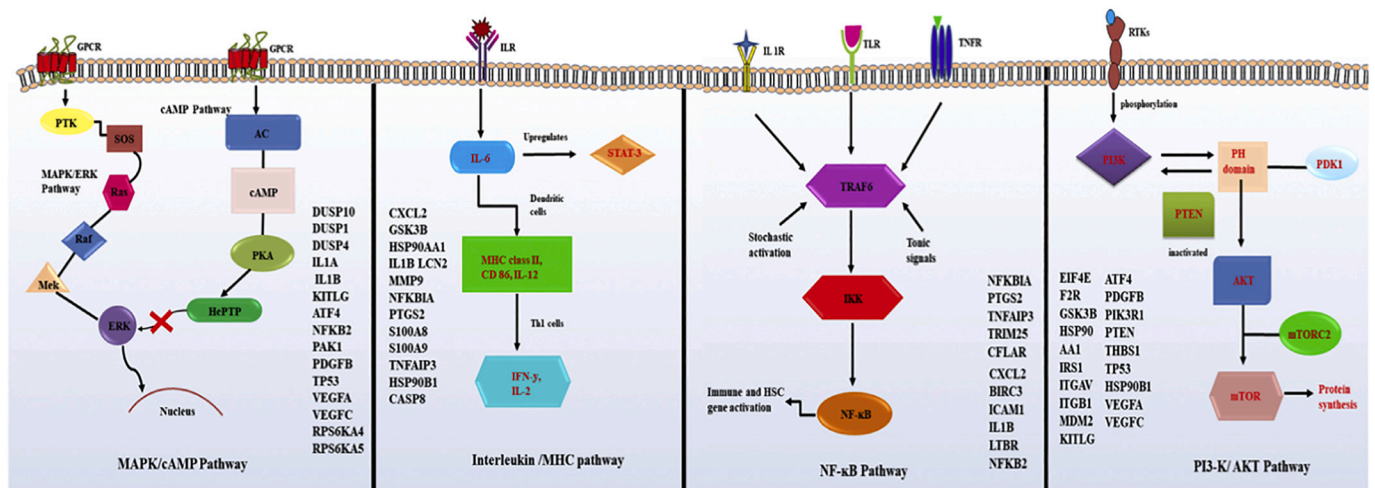
Majority of differentially expressed genes were belonging to NF-kappa B pathway, PI3K-AKT pathway, MAPK/ERK pathway, cAMP pathway, and interleukin pathway, and MHC class 1 mediated antigen processing pathway (Fig. 1). NF-kappa B pathway acts as an enhancer for B cell activation. This pathway controls the transcription of DNA, cytokine production, cell survival [41]. This pathway is prevalent in all cell types and helps in regulation of immune response against infections. NF-kappa B activation involves canonical as well as non-canonical pathway which is important in regulation of immune and inflammatory response [42]. It includes ligand for various cytokine receptors, pattern recognition receptors (PRRs) and T cell, B cell receptors. The innate immune cells express PRRs which detect microbial components that is, pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs) [43]. PI3K/AKT is an intracellular signal transduction pathway which controls the metabolism, proliferation, cell survival, and angiogenesis in presence of an external signal [44]. They are mediated via serine/threonine phosphorylation. The major proteins involved in this pathway are phosphatidylinositol 3-kinase (PI3K) and AKT/Protein kinase B [45]. AKT regulates the function of various proteins via phosphorylation which promotes cell growth and proliferation. AKT negatively regulates the Bcl-2 members and prevents apoptosis. Over activation of this pathway leads to abnormal cell proliferation which leads to tumorigenesis [46]. MAPK/ERK pathway is also called Ras-Ref-MEK-ERK, it's a chain of protein that passes on the signal from the receptor to the DNA in the nucleus in cells [47]. GPCR leads to the activation MAPKs by phosphorylation of receptor tyrosine kinase such as epidermal growth factor receptor (EGFR) gets activated in presence of extracellular ligand [48]. This pathway regulates cell growth and proliferation in mammalian cells. The canonical pathway is triggered in presence of mitogen and other growth factors which leads to the activation of GTPase Ras, followed by the series of phosphorylation and activation of MAPK cascade which ultimately leads to the activation of ERK. ERK promotes gene activation that induces cell cycle entry and downregulates the negative regulators of cell cycle, loss of control can lead to cancer [49]. cAMP pathway is a G protein coupled receptor relates signalling cascade. cAMP regulates metabolism, secretion, calcium homeostasis and gene transcription [50]. cAMP acts as secondary messenger in various

**Table 1**

List of top 10 up and down regulated genes in SARS-CoV-2 infected NHBE cell lines.

Genes	P value	Fold change	TPM (Mock)	TPM (Infection)	Description
ENSG00000121858	2.95E-07	-3.252	73.471	9.201	Tumor necrosis factor (ligand) superfamily
ENSG00000075624	0.0018	-3.034	5489.099	1796.564	Actin
ENSG00000142192	0.0002	-2.832	210.277	15.893	Amyloid beta (A4) precursor protein
ENSG00000162772	4.19E-06	-2.763	24.919	5.715	Activating transcription factor 3
ENSG00000141367	0.0005	-2.461	63.660	14.057	Clathrin
ENSG00000078142	0.0009	-2.191	13.849	5.136	Phosphatidylinositol 3-kinase
ENSG00000125538	0.0013	-2.052	149.285	53.990	Interleukin 1
ENSG00000108599	0.0003	-2.031	30.403	8.495	A kinase (PRKA) anchor protein 10
ENSG00000179750	0.0005	-1.986	25.141	6.919	Apolipoprotein B mRNA editing enzyme
ENSG00000015171	0.0021	-1.737	80.428	21.587	Zinc finger
ENSG00000157601	1.64E-08	3.607	11.958	120.264	Myxovirus (influenza virus) resistance 1
ENSG00000105372	0.0029	3.165	1576.726	3727.784	Ribosomal protein S19
ENSG00000100979	2.38E-05	3.098	3.235	52.695	Phospholipid transfer protein
ENSG00000170581	4.10E-06	2.912	6.547	53.179	Signal transducer and activator of transcription 2
ENSG00000160014	7.21E-05	2.904	78.825	338.738	Calmodulin 3 (phosphorylase kinase)
ENSG00000165949	8.43E-07	2.840	12.392	128.770	Interferon
ENSG00000080503	0.0005	2.765	18.355	41.172	SWI/SNF related
ENSG00000090339	1.15E-06	2.726	27.981	174.443	Intercellular adhesion molecule 1
ENSG00000145901	0.0014	2.523	56.837	254.358	TNFAIP3 interacting protein 1
ENSG00000160710	0.0008	2.450	6.969	26.304	Adenosine deaminase
ENSG00000135018	1.53E-05	2.335	32.631	158.313	Ubiquitin 1





**Fig. 1.** Schematic representation of the pathways expressed during humoral immune response of SARS-CoV-2 infection. A) MAPK/cAMP pathway; B) Interleukin/MHC pathway; C) NF-kappa B pathway; D) PI3K-AKT signal transduction pathway. Camp acts an intracellular regulator of cellular homeostasis and also negatively regulates proinflammatory signalling, cytokines such as TNF- $\alpha$  and IL-1 $\beta$  leads to decrease in cAMP in the cells, but mechanism behind this action is poorly known so far. In presence of TNF- $\alpha$ , ERK 1/2 is phosphorylated which is present in the MAPK signalling cascade, NF-Kappa B p65 gets translocated in the nucleus all these occurs when there are lower levels of cAMP [55]. The MAPK pathway and PI3K-AKT pathway are two independent signalling pathways which plays important role in cell survival, proliferation and motility. These pathways also cross talk extensively to either positively or negatively regulate each other [56].

signalling pathways. AC is regulated by G-alpha which induces conversion of ATP into cAMP [51]. cAMP has three major targets protein kinase A (PKA), exchange protein activation and cyclic nucleotide gated ion channels. This pathway activates enzyme and regulation of genes. If the cAMP dependent pathways aren't controlled leads to hyper proliferation which leads to progression of cancer [52]. Interleukin family of cytokines are composed of 11 proteins encoded by 11 different genes. They act as major mediators of innate immune response. They induce a cascade of proinflammatory cytokines *via* the expression of integrins on leukocytes and endothelial cells, which in turn regulates and controls inflammatory response [53]. IL-6 acts a proinflammatory cytokine and an anti-inflammatory myokine. IL-6 is an important mediator of fever and of acute phase response. They are secreted by macrophages in presence of microbial components referred to as PAMPs [53]. PAMPs bind to PRRs and Toll like receptors which induces intracellular signalling cascade, which is responsible for inflammatory cytokine production. As a stress response our body produces stress hormones like cortisol which triggers IL-6 and releases it in circulation. There is some early evidence that IL-6 can be used as an inflammatory marker for severe COVID-19 infection with poor prognosis [54]. MAPK/ERK pathway upregulates the production of epidermal growth factors and down regulates the negative regulators of cell cycle whereas interleukin/MHC pathway regulates the level of cytokines which can mediate the innate immune response. NF-kappa B pathway acts as an enhancer for the B cell activation and PI3K-AKT a signal transduction pathway which is known to controls the proliferation and survival of the cells.

### 3.3. SARS-CoV-2 transcriptome assembly and annotation

*De novo* SARS-CoV-2 transcriptome was constructed through high quality of reads from cell line transcriptome. Total, 87,716 reads were extracted from all the samples for SARS-CoV-2 genome. Detail description of experiment, sample name, description was given in Supplementary File 1, Table S1. All the extracted RNA-seq reads were used to construct *de novo* SARS-CoV-2 transcriptome through Trinity software. In total, 54,814 bases were assembled into 27 transcripts with median contig length 650 bps, N50 value of 10,677 bps and approximate average transcript length of 2030 bps. The generated transcriptome assembly was clustered at 90% sequence identity through CD-HIT software that produced 27 non-redundant transcripts, the same

number of non-redundant transcripts showed the sequence variability among assembled transcripts. Non-redundant transcripts were translated into 44 protein sequences, and annotated against NCBI non-redundant and Uniport databases by using BLAST similarity search. All the reads were mapped over constructed transcriptome by using Kallisto software to estimate expression of transcripts. Detail description of transcripts class, annotation and expression was given in Table 2.

### 3.4. T-cell epitopes identification of SARS-CoV-2 transcriptome

T-cell epitopes are presented by MHC class I and II that are recognized by two distinct subsets of T-cells, CD8+ and CD4+ T-cells, respectively. NetCTL 1.2 program was used for the prediction of T-cell epitopes, and 1144 epitopes were selected for 12 super type categories *i.e.* A1 (330), A24(314), A26(242), A2(247), A3(328), B27(175), B39(263), B44(133), B58(284), B62(473), B7(157), and B8(193). On the basis of antigenicity and immunogenicity potential, 598 epitopes were selected for further study. Peptide length nine and the IC50 value  $\leq 200$  nM parameters were selected to identify MHC alleles and genes, and ToxinPred was used to explore toxicity of predicted epitopes. Finally, 40 CD8+ T-cell epitopes were selected as high immunogenic, antigenic, non-toxic epitopes with good binding affinity to MHC class I alleles (Supplementary File 2, Table S1). In MHC-I allele analysis, HLA-A type alleles were found as more frequently occurring alleles (HLA-A\*01:01, HLA-A\*02:01, HLA-A\*02:03, HLA-A\*02:06, HLA-A\*03:01, HLA-A\*11:01, HLA-A\*23:01, HLA-A\*24:02, HLA-A\*26:01, HLA-A\*30:01, HLA-A\*30:02, HLA-A\*31:01, HLA-A\*32:01, HLA-A\*33:01, HLA-A\*68:01, HLA-A\*68:02) than HLA-B type alleles (HLA-B\*07:02, HLA-B\*08:01, HLA-B\*15:01, HLA-B\*35:01, HLA-B\*40:01, HLA-B\*44:02, HLA-B\*44:03, HLA-B\*53:01, HLA-B\*57:01).

Similarly, MHC class II gene and allele's prediction was performed through IEDB analysis resources by using SMM method for the prediction of a distinct set of MHC class II alleles. Total, 4072 epitopes were identified from protein sequences with good binding affinity to MHC class II alleles. To select MHC class II alleles and epitopes, we decided to take those MHC class II alleles and epitopes which have MHC class I epitopes peptide sequence as a core sequence. Total, 34 MHC class II epitope sequences (15-mer) were selected by using previously selected 40 (9-mer) antigenic and immunogenic epitope sequences as core sequences (Supplementary File 2, Table S2). Among all MHC class II alleles, HLA-DPA1\*01:03/DPB1\*04:01,

HLA-DPA1\*02:01/DPB1\*14:01, HLA-DRB3\*02:02, HLA-DRB1\*01:01, HLA-DRB1\*07:01, HLA-DPA1\*01:03/DPB1\*02:01, and HLA-DRB5\*01:01 were the most frequent occurring alleles. Finally, twelve most antigenic and immunogenic MHC class II epitopes (APHGVVFLHVTYVPA, FLHVTYVPAQEKNFT, GVVFLHVTYVPAQEK, HGVVFLHVTYVPAQE, PHGVVFLHVTYVPAQ, QSAPHGVVFLHVTYV, QYIKWPWYIWLGFIA, SAPHGVVFLHVTYV, VFLHVTYVPAQEKNF, VVFLHVTYVPAQEKN, and YIKWPWYIWLGFIA) were selected which contains 9-mer core sequences of four epitopes from previously selected 40 epitopes (Table 3).

Genetic variability among MHC alleles is a major obstacles in the development of peptide-based vaccines. Therefore, population coverage is an important criterion to design a generalized an effective vaccine [57]. In population coverage analysis, MHC class I allele's of 40 epitopes and MHC class II allele's of 33 epitopes were used, and a significant population coverage was found for different geographic regions around the world. MHC class alleles of selected epitopes were covered approximately 90% of the world population. Highest population coverage was found for Sweden (100%) which was closely followed by England, Germany, France, Belgium, United States, Russia, Italy, South Korea, Japan, Mexico, Iran, Chile, Brazil, China, Singapore, Pakistan, India, Spain, Thailand, Israel, Philippines, Australia, and Vietnam with a population coverage of 99.99%, 99.99%, 99.97%, 99.87%, 99.87%,

99.81%, 99.78%, 99.66%, 99.55%, 99.02%, 98.84%, 98.55%, 98.43%, 97.6%, 97.24%, 97.13%, 97.05%, 96.9%, 96.48%, 96.47%, 96.41%, 96.17%, and 95.91% respectively. The lowest population coverage were found for Canada (38.31%), Sri Lanka (42.04) and Ukraine (46.48). United States and Europe has highest number of COVID-19 cases [58]. Hence, the population coverage prediction is essential for vaccine design. Population of ethnic groups were also significantly covered (Supplementary File 2, Table S3), and average coverage for ethnic group across the world is around 93%.

### 3.5. B-cell epitope identification of SARS-CoV-2 transcriptome

B-cell epitope is a precise region of the antigenic protein that is detected by B-cell receptors (BCR) through membrane-bound immunoglobulins. Once B-cell activated, it secretes soluble forms of the immunoglobulins to neutralize antigenic proteins. Thus, B-cell epitope and B-cell receptor information is essential for epitope-based vaccine design. Prediction of B-cell epitopes was performed by using protein sequences of the assembled transcriptome and SARS-CoV-2 protein structures. Total, 330 B-cells epitopes were predicted through BepiPred-2.0 program by using protein sequences whereas 77 B-cell epitopes were predicted through the IEDB conformational tool ElliPro by using

**Table 2**

SARS-CoV-2 transcriptome annotation through BLAST similarity search at e-value threshold 1e-10 along with their expression values.

Transcripts ids	Functional class	Uniport annotation	Expression (TPM)
TRINITY_DN10_c0_g1_i1_p1	H	ORF1ab polyprotein Tax = BtRs-BetaCoV/YN2013 TaxID = 1,503,303	3661.09
TRINITY_DN0_c0_g1_i1_p3	M	Membrane protein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	91,746.3
TRINITY_DN0_c0_g1_i2_p5	M	Membrane protein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	1716.75
TRINITY_DN0_c0_g1_i4_p5	M	Membrane protein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	3188.07
TRINITY_DN0_c0_g1_i6_p5	M	Membrane protein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	12,432
TRINITY_DN0_c0_g1_i1_p1	N	Nucleoprotein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	91,746.3
TRINITY_DN0_c0_g1_i2_p3	N	Nucleoprotein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	1716.75
TRINITY_DN0_c0_g1_i4_p3	N	Nucleoprotein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	3188.07
TRINITY_DN0_c0_g1_i6_p3	N	Nucleoprotein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	12,432
TRINITY_DN0_c0_g1_i7_p1	N	Nucleoprotein Tax = Bat SARS-like coronavirus TaxID = 1,508,227	838,657
TRINITY_DN0_c0_g1_i2_p2	NendoU	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	1716.75
TRINITY_DN0_c0_g1_i4_p2	NendoU	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	3188.07
TRINITY_DN0_c0_g1_i6_p2	NendoU	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	12,432
TRINITY_DN0_c0_g1_i3_p1	NSP1	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	10,488.1
TRINITY_DN0_c0_g1_i5_p1	NSP1	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	3765.1
TRINITY_DN20_c0_g1_i1_p1	NSP2	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	1786.89
TRINITY_DN4_c0_g1_i1_p1	NSP2	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	3107.15
TRINITY_DN13_c0_g1_i1_p1	NSP3	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	1009.35
TRINITY_DN17_c0_g1_i1_p1	NSP3	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	1416.36
TRINITY_DN19_c0_g1_i1_p1	NSP3	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	1759.06
TRINITY_DN2_c0_g1_i1_p1	NSP3	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	1602.68
TRINITY_DN5_c0_g1_i1_p1	NSP3	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	553.586
TRINITY_DN6_c0_g1_i1_p1	NSP3	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	1813.9
TRINITY_DN7_c0_g1_i1_p1	NSP6	Non-structural polyprotein 1ab Tax = Betacoronavirus TaxID = 694,002	697.846
TRINITY_DN1_c0_g1_i1_p1	NSP8	Non-structural polyprotein 1ab Tax = Bat SARS-like coronavirus	1941.65
TRINITY_DN0_c0_g1_i1_p2	ORF3a	Uncharacterized protein Tax = Human SARS coronavirus TaxID = 694,009	91,746.3
TRINITY_DN0_c0_g1_i2_p4	ORF3a	Uncharacterized protein Tax = Human SARS coronavirus TaxID = 694,009	1716.75
TRINITY_DN0_c0_g1_i4_p4	ORF3a	Uncharacterized protein Tax = Human SARS coronavirus TaxID = 694,009	3188.07
TRINITY_DN0_c0_g1_i6_p4	ORF3a	Uncharacterized protein Tax = Human SARS coronavirus TaxID = 694,009	12,432
TRINITY_DN0_c0_g1_i1_p4	ORF7a	SARS_X4 domain-containing protein Tax = Bat SARS-like coronavirus	91,746.3
TRINITY_DN0_c0_g1_i2_p6	ORF7a	SARS_X4 domain-containing protein Tax = Bat SARS-like coronavirus	1716.75
TRINITY_DN0_c0_g1_i4_p6	ORF7a	SARS_X4 domain-containing protein Tax = Bat SARS-like coronavirus	3188.07
TRINITY_DN0_c0_g1_i6_p6	ORF7a	SARS_X4 domain-containing protein Tax = Bat SARS-like coronavirus	12,432
TRINITY_DN0_c0_g1_i1_p5	ORF8	Uncharacterized protein, Bat SARS-like coronavirus TaxID = 1,508,227	91,746.3
TRINITY_DN0_c0_g1_i2_p7	ORF8	Uncharacterized protein, Bat SARS-like coronavirus TaxID = 1,508,227	1716.75
TRINITY_DN0_c0_g1_i4_p7	ORF8	Uncharacterized protein, Bat SARS-like coronavirus TaxID = 1,508,227	3188.07
TRINITY_DN0_c0_g1_i6_p7	ORF8	Uncharacterized protein, Bat SARS-like coronavirus TaxID = 1,508,227	12,432
TRINITY_DN15_c0_g1_i1_p1	Proteinase	UPI0001D192D5 related cluster, TaxID = RepID = UPI0001D192D5	1706.65
TRINITY_DN9_c0_g1_i1_p1	Proteinase	UPI000181CE36 related cluster, TaxID = RepID = UPI000181CE36	811.289
TRINITY_DN11_c0_g1_i1_p1	RdRp	ORF1ab polyprotein n = 1 Tax = BtRf-BetaCoV/JL2012 TaxID = 1,503,299	1168.03
TRINITY_DN12_c0_g1_i1_p1	RdRp	RNA-dependent RNA polymerase Tax = Human SARS coronavirus	1833.88
TRINITY_DN0_c0_g1_i2_p1	SG	Spike protein n = 1 Tax = Bat SARS-like coronavirus TaxID = 1,508,227	1716.75
TRINITY_DN0_c0_g1_i4_p1	SG	Spike protein n = 1 Tax = Bat SARS-like coronavirus TaxID = 1,508,227	3188.07
TRINITY_DN0_c0_g1_i6_p1	SG	Spike protein n = 1 Tax = Bat SARS-like coronavirus TaxID = 1,508,227	12,432

H: Helicase, M: Membrane protein, N: Nucleoprotein, NendoU: Uridylate-specific endoribonuclease, NSP: Non-structural protein, ORF: open reading frame, SG: Surface glycoprotein, RdRp: RNA-dependent RNA polymerase, TPM: Transcript per million.

**Table 3**  
MHC class II alleles and epitopes along with their core 9-mer epitopes.

Anno	MHC class I epitope sequences	Immscore	Antiscore	MHC class II alleles	MHC class II epitope sequences
SG	QYIKWPWYI WPWYIWLGF	0.41673	1.4953	HLA-DPA1*01:03/DPB1*02:01 HLA-DPA1*01:03/DPB1*04:01	QYIKWPWYIWLGFIA YIKWPWYIWLGFIA
SG	VVFLHVTYV	0.1278	1.5122	HLA-DRB1*01:01; HLA-DRB1*04:05; HLA-DRB1*07:01; HLA-DRB3*02:02;	QSAPHG VVFLHVTYV SAPHG VVFLHVTYV APHG VVFLHVTYVPA PHG VVFLHVTYVPAQ
SG	FLHVTYVPA	0.11472	1.3346	HLA-DPA1*01:03/DPB1*02:01; HLA-DPA1*01:03/DPB1*04:01; HLA-DPA1*02:01/DPB1*14:01; HLA-DRB1*01:01; HLA-DRB1*04:05; HLA-DRB1*07:01; HLA-DRB3*02:02;	APHG VVFLHVTYVPA PHG VVFLHVTYVPAQ HG VVFLHVTYVPAQE GVVFLHVTYVPAQEK VVFLHVTYVPAQEK VFLHVTYVPAQEKNF FLHVTYVPAQEKNF

Immscore: Immunogenicity score, Antiscore: Antigenicity score, SG: Surface glycoprotein.

24 SARS-CoV-2 protein structure. B-cell epitopes prediction from the protein structure is highly useful information for epitope-based vaccine design and development [59,60]. Therefore, a separate analysis was performed for B-cell epitopes identified from protein structures (Supplementary File 2, Table S4). In order to explore most suitable B-cell epitopes, epitopes generated by both the approaches were combined, and 16 non-toxic, immunogenic and antigenic B-cell epitopes were identified. A cross-reactivity analysis of epitopes were performed against human proteome sequences through BLAST similarity search, and found that two B-cell epitopes (DNNFCGPDGYPLE, NQDLNGN-WYD) showed significant similarity with human proteins ENSP00000390696.1 and ENSP00000263390.3 respectively. Finally, 14 epitopes were selected as most potential B-cell epitopes for further analysis (Table 4).

On the basis of immunological parameters, the five B-cell epitopes, NSP2 (SEQLDFIDTKRGV, HCGETSWQTGDFV), Helicase (KGTLEPEYF), NSP3 Papain-like (KTVGELGDVRE), and Surface glycoprotein (LTGT-GVLTESNK) were selected from Table 4 for molecular docking studies of B-cell receptors.

### 3.6. Molecular docking analysis of T- and B cell epitopes

Cellular immunity gets activated when MHC molecules binds to intracellular and extracellular proteins displayed on the cell surface. Therefore, molecular docking studies of epitopes and MHC molecules can enhance our knowledge about molecular binding of peptide and receptors. Docking studies were performed between MHC protein and T-cell epitopes to evaluate binding affinities of epitopes. Twenty-two

MHC proteins were explored with selected three T-cell epitopes, and the best interaction was identified with the highest binding affinity. The docked complexes of epitopes (WPWYIWLGF, VVFLHVTYV, and FLHVTYVPA) and the HLA molecules were retaining a binding affinity range from of  $-136.54$  to  $-7.12$  kcal/mol. Detail description of molecular interaction analysis of peptide VVFLHVTYV and identified protein structure of MHC genes were given in Table 5. Detail docking descriptions of other two epitopes with MHC molecules were given in Supplementary File 2, Table S5.

HLA-A\*24:02 allele was highly expressed among most frequent occurring MHC alleles, but highest interaction of peptide (VVFLHVTYV) was shown with HLA-DRB3\*02:02 allele ( $-1560.1$  kcal) and contained two hydrogen bonds (THR7-ASP152, LEU4-SER120). The epitope position in protein structure and the binding interactions were shown in Fig. 2. We were also ensured genuine hydrogen bonds interaction with receptors through cavity prediction (Supplementary File 3).

Similarly, five most antigenic and immunogenic B-cell epitopes were selected for molecular docking studies with two B-cell receptors (5DRW and 1K1F). 5DRW protein structure is a crystal structure of BCR Fab fragment from subset of chronic lymphocytic leukaemia whereas 1K1F is dimer structure of Bcr-Abl oncoprotein. 1K1F protein structure provided a base to design an inhibitor to disrupt Bcr-Abl oligomerization. Moreover, 1K1F was without Fab region unlike to 5DRW [61]. In our analysis, most of the peptides were shown higher binding affinities with 5DRW than 1K1F. Detail description of docking study of B-cell receptor and peptides are given in Table 6.

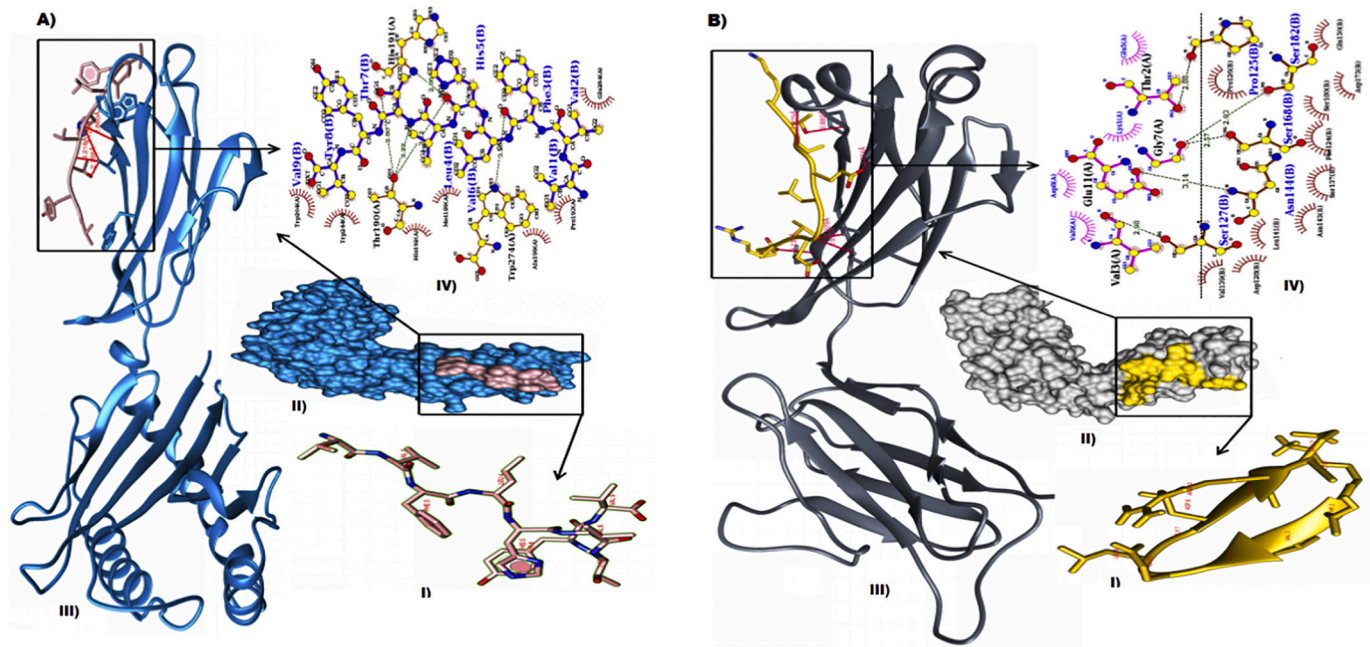
For protein structure 5DRW, interaction and total energy of peptide and B-cell receptor were varied from  $-94.17$  to  $-37.47$  kcal/mol

**Table 4**  
B-cell epitopes identified form protein sequences of assembled transcriptome and modelled protein structure of various coding protein of SAR-CoV-2 genome.

B-cell Epitopes	Anno	Len	Start	End	Immscore	Antiscore	Toxicity
SEQLDFIDTKRGV	NSP2	13	36	48	0.13632	1.7773	NT
KGTLEPEYF	H	9	414	422	0.17084	1.3504	NT
HCGETSWQTGDFV	NSP2	13	145	157	0.24817	1.1314	NT
KTVGELGDVRE	NSP3	11	902	912	0.25316	0.9231	NT
LTGTGVLTESNK	SG	12	546	557	0.10111	0.8122	NT
TGVVGESEGLN	NSP2	12	256	267	0.17371	0.7539	NT
QTTETAHSC	H	9	548	556	0.11862	0.7078	NT
MEVTPSGTWLT	N	11	322	332	0.13056	0.5982	NT
SDARTAPHG	NSP1	9	74	82	0.14895	0.5706	NT
LKATEETFK	H	9	138	146	0.34467	0.5278	NT
NENGTITDA	SG	9	280	288	0.2408	0.5257	NT
KGHFDGQQGEVPS	NendoU	14	12	25	0.14142	0.5183	NT
LQAGNATEVPANS	NSP10	13	1	11	0.2729	0.4491	NT
VQIPTTCANDPVGFT	NSP10	15	97	111	0.2336	0.4488	NT

Immscore: Immunogenicity score, Antiscore: Antigenicity score, NT: Non-toxic, H: Helicase, M: Membrane protein, N: Nucleoprotein, NendoU: Uridylate-specific endoribonuclease, NSP: Non-structural protein, ORF: open reading frame, SG: Surface glycoprotein, RdRp: RNA-dependent RNA polymerase, 2'-O-MT: 2'-O-methyltransferase, ExoN: Guanine-N7 methyltransferase.





**Fig. 2.** Docking of between peptide and protein receptor. A) T-cell peptide (VVFLHVTYV) and HLA allele (HLA-A\*24:02) protein structure (5XOV); B) B-cell peptide (KTVGELGDVRE) and protein structure of B-cell receptor (5DRW); I) Epitope structure; II and II) Cartoon structure of docked molecule (hydrogen bonds in red colour); d) Molecular level description of interaction through LIGPlot software (hydrogen bonds in green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and  $-1656.88$  to  $-1235.01$  kcal/mol respectively. Whereas interaction and total energy for 1K1F were varied from  $-110.1$  to  $-25$  kcal/mol and  $-686.76$  to  $-362.98$  kcal/mol respectively (Supplementary File 2, Table S6).

#### 4. Discussion

Contagious nature of SARS-CoV-2 virus, infected more than thirteen million people worldwide has imposed the biggest challenge of COVID-19 treatment and prevention. In order to develop epitope-based vaccine, the surface glycoprotein is the primary interest because it is

involved in the interaction between virus and human cell receptor. But, other viral elements are also important to cause disease [62]. Information of expressed regions of virus genome is very important to identify potential vaccine candidates. In the present study, MHC gene alleles and innate immune response were analysed through transcriptome data of cell lines, and SARS-CoV-2 transcriptome was generated for the molecular cataloguing of immunodominant epitopes. The result of performed analysis is summarized in Fig. 3. In this study, SARS-CoV-2 transcriptome was constructed through RNA-seq data of SARS-CoV-2 infection in human cell lines NHBE and A549 to explore expressed regions of SARS-CoV-2 genome. In assembled transcripts,

**Table 5**

Docking analysis of T-cell epitope (VVFLHVTYV) and MHC protein structures.

MHC alleles (PDB code)	Expression (TPM)	Protein chain	Interaction energy	Total energy	Hydrogen bond (Peptide – Receptor)
HLA-DRB3*02:02 (2Q6W)	NoExp	B	-96.73	-1560.1	THR7-ASP152, LEU4-SER120
HLA-A*02:03 (3OX8)	46.1625	A	-63.72	-1937.21	THR7-ASP77, THR7-TYR84
HLA-DQA1*05:01/DQB1*03:01 (4D8P)	NoExp	A	-69.82	-890.23	VAL1-SER8, TYR8-THR93, VAL1-VAL6, PHE3-SER8, TYR8-THR83, TYR8-ASP142
HLA-DRB1*04:01 (5JLZ)	NoExp	B	-57.25	-1377.11	THR7-GLU187, HIS5-GLU187, PHE3-VAL101
HLA-DRB1*01:01 (5V4N)	NoExp	C	-63.35	-1430.47	VAL1-HIS360, THR7-GLU411
HLA-A*23:01 (5WVJ)	78.8768	C	-95.26	-1611.15	TYR8-CYS264, TYR8-CYS388, VAL9-TYR327, VAL9-ASN331
HLA-A*24:02 (5XOV)	2597.88	A	-82.36	-2186.76	THR90-HIS5, HIS191-HIS5, THR190-THR7, THR190-THR7
HLA-A*11:01 (6ID4)	NoExp	A	-7.12	-1452.3	LEU4-TYR27, LEU4-ARG6, TYR8-SER4, SER4-VAL9
HLA-B*44:02 (3DX6)	NoExp	A	-83.81	-2384.29	THR7-ASP114, VAL1-SER167, VAL1-GLU55
HLA-DPA1*01:03/DPB1*02:01 (3LQZ)	NoExp	A	-75.51	-3195.72	HIS5-GLU134, HIS5-HIS144
HLA-A*02:06 (3OXR)	NoExp	C	-47.82	-2993.06	VAL1-HGLN54, PHE3-GLN54, HIS5-GLU55, HIS5-TRP51
HLA-B*57:01 (3X11)	NoExp	A	-31.02	-2183.55	TYR8-GLN155, VAL9-TYR99, VAL1-SER116, PHE3-SER8
HLA-A*68:02 (4HWZ)	61.582	A	-56.06	-2054.65	PHE3-ARG144, VAL1-ASP77, VAL9-THR73, HIS5-THR73
HLA-B*44:03 (4JQX)	38.0344	A	-39.35	-1709.95	THR7-ASN77, VAL9-SER69, TYR6-GLN89
HLA-B*35:01 (4PRA)	53.433	A	-22.74	-1780.49	VAL1-ARG6, VAL9-SER116, THR7-ASP144, TYR8-TYR74
HLA-A*02:01(4U6Y)	27.0879	A	-29.1	-1507.37	VAL1-GLU63, VAL1-TYR99, HIS5-HIS114, THR7-HIS144
HLA-DRB1*11:01(5NI9)	NoExp	B	-90.52	-1566.14	VAL6-TYR30, TYR8-TYR48
HLA-B*58:01 (5VWH)	NoExp	A	-76.23	-2288.87	THR7-ARG14, HIS5-ARG21
HLA-A*30:02 (6JIV)	NoExp	A	-80.19	-2377.92	VAL6-ARG202
HLA-A*30:01 (6J1W)	NoExp	A	-46.93	-1869.26	PHE3-ARG273, HIS5-THR271
HLA-A*03:01 (6O9B)	23.4843	A	-62.78	-2263.52	TYR8-TYR-123, TYR8-ARG114, HIS5-ARG114
HLA-A*68:01 (6PBH)	61.582	A	-88.24	-2076.8	TYR6-TYR99, ALA9-THR143

TPM (Transcript per million).



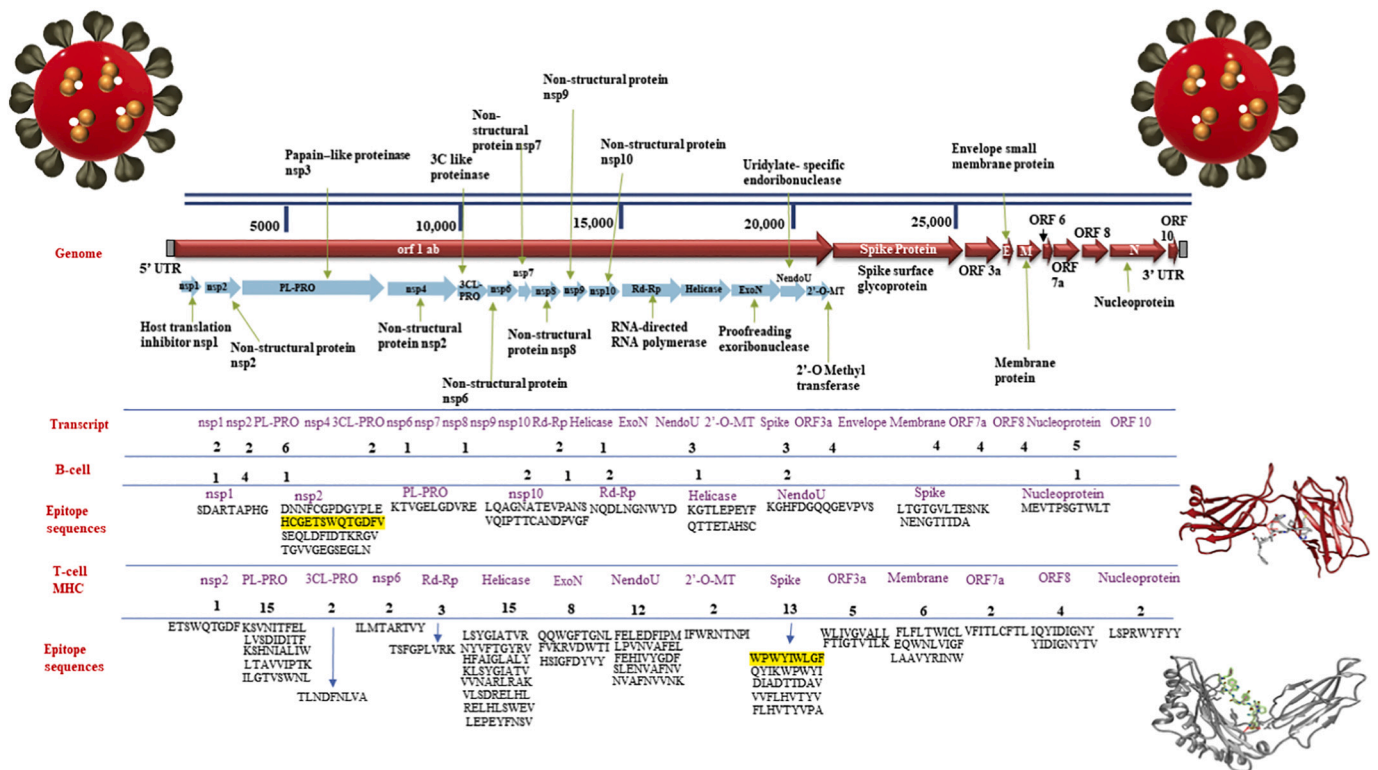
**Table 6**  
Molecular docking and interaction analysis of B-cell epitope to B-cell protein receptor.

Peptide	PDB code	Protein chain	Interaction energy	Total energy	Hydrogen bonds (Peptide – Receptor)
SEQLDFIDTKRGV	5DRW	B	-94.17	-1656.88	THR9-PRO49, THR9-SER48, THR9-ARG51, GLU7-LYS55
	1K1F	A	-25.53	-362.98	THR9-ARG50, TRP7-ARG51, GLU4ARG51, CYS2-ASN39
HCGETSWQTGDFV	5DRW	B	-85.9	-1579.94	SER6-GLN205, THR9-GLN205
	1K1F	A	-29.14	-366.08	GLN8-ARG22
KGTLEPEYF	5DRW	B	-91.21	-1767.94	THR3-SER20, GLU7-SER10
	1K1F	A	-35.76	-629.66	GLU5-GLN51
KTVGELGDVRE	5DRW	B	-37.47	-1490.12	THR2-PRO125, VAL3-SER127, GLU11-ASN144, GLY7-SER168, GLY7-SER182
	1K1F	A	-54.36	-686.76	THR5-ARG25
LTGTGVLTESNK	5DRW	B	-40.46	-1235.01	THR4-TYR37, THR4-ASN39, SER10-GLN43
	1K1F	A	-110.1	-652.47	GLU9-SER18, GLU9-SER41

non-structural polyprotein 1ab related transcripts were found as most abundant transcripts, whereas nucleoprotein, membrane protein, SARS\_X4 domain-containing protein, spike protein were also found in significant numbers. Targeted expression profiling of MHCs gene and alleles was performed through two different kind of cell line data to explore expression variation between cell lines against the infection. To overcome the effects of reads fail to capture and ambiguously mapping reads, differential gene expression analysis was performed between mock and infected samples of NHBE cell line through Kallisto-Sleuth pipeline [26]. Workflow of study was given in Supplementary File 3.

The largest region of SARS-CoV-2 genome (ORF1ab) was expressed various smaller proteins such as non-structural proteins (NSP1, NSP2, NSP3, NSP4, NSP6 and NSP8), helicase, uridylylate-specific endoribonuclease, RNA-dependent RNA polymerase. Expression of polyprotein 1ab subcomponent (Table 2) is clearly reflecting about disease initiation and progression such as expression of NSP1 protein was indicating the inhibition of host translation machinery by making NSP1-40S ribosome complex which can cause an endonucleolytic cleavage near the 5' UTR of host mRNAs for degradation. NSP1 also facilitates

viral gene expression in infected cells by suppressing host gene expression [63]. Role of EIF4E gene is very important for cellular growth control because it initiates protein synthesis on capped mRNAs in the cytoplasm [64]. In our DGE analysis, ENST00000450253 and ENST00000505992 transcripts of eukaryotic translation initiation factor 4E (EIF4E) were found as down regulated in infected samples which might influence the ribosomal protein turnover in cytoplasm. Reduced protein turnover in cytoplasm was reflected by down regulation of RPS27A and RPS19 genes in our analysis. Two transcript of RPS19 gene (ENSG00000105372) were down regulated whereas third transcript (ENST00000221975) was three fold up regulated. The up regulation of ENST00000221975 might be because of viral mRNA translation for disease initiation. NSP2 is another expressed transcripts that may have a role in the alteration of host cell survival signalling pathway by interacting with host prohibitin (PHB) and prohibitin-2(PHB2) [65]. Gene FOXO1 (ENSG00000150907), GSK3B (ENSG00000082701), TRIM14 (ENSG00000106785), VNN1 (ENSG00000112299) and PIK3C3 (ENSG00000078142) genes were found related to cell survival AKT-signalling pathway in host cell. Acute



**Fig. 3.** Genome-wide transcriptome analysis of SARS-CoV-2, T- and B-cell epitopes identification for multi-peptide based vaccine development. Docked structures of B-cell receptor and MHC alleles are given at right-side and their T- and B-cell epitopes are highlighted in yellow colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

inflammatory response (VNN1) and negative regulation of viral transcription associated genes (TRIM14) were up regulated to elite primary immune response.

Papain-like proteinase (NSP3) is responsible for N-terminus cleavage of the replicase polyprotein and involved in the assembly of virally-induced cytoplasmic double-membrane vesicles together with NSP4, necessary for viral replication. NSP3 is an important viral molecular factor to suppress innate immune induction of type I interferon by blocking the phosphorylation, dimerization and subsequent nuclear translocation of host interferon regulatory transcription factor (IRF3), and also suppress NF-kappa-B signalling of host [66,67]. In DGE analysis, Gene (ENSG00000119922) encodes interferon induce protein in response to virus and interleukin 1 (ENSG00000125538) were down regulated, whereas IFI27, IFI6, IFITM1, IFITM3, IRF9, IL32, IL8, IRAK2, and IRAK3 were up regulated. It has been demonstrated that SARS-CoV-2 infection fails to induce significant type I interferon responses in blood but induces the subsequent upregulation of IFN-stimulated genes [68,69]. NSP6 initiate induction of autophagosomes from host reticulum endoplasmic. But later, it reduces the expansion of phagosomes to stop delivery of viral components to lysosomes [70]. Transcription factor EB, E2 transcription factor, and forkhead box O genes are involved in autophagosome biogenesis and upregulation of target genes such as LC3, cathepsin D, and LAMP1 to the increasing activity of autophagy-lysosome pathway (ALP). Genes associated with ALP, FOXO1 (ENSG00000150907), cathepsin D (ENSG00000117984) and cathepsin H (ENSG00000103811) were found down regulated and down regulation of these genes might be the reason of reduce delivery of viral components to lysosomes. NSP8, together with NSP7 forms hexadecamer act as a primase to participate in a viral replication. Expression of helicase is required for the RNA and DNA duplex-unwinding activities [71].

Constructed SARS-CoV-2 transcriptome sequences were used to identify potential epitopes for T- and B-cells. Best immunodominant epitopes were identified after the careful filtering with different parameters such as antigenicity ( $\geq 1$ ) and immunogenicity ( $\geq 0.1$ ), IC50 values ( $IC_{50} = <200$ ), toxicity, population coverage and conservation. Antigenicity of epitopes represents the ability to bind or interact with B-cell or T-cell receptors, whereas immunogenic features of epitopes triggers the innate immune response to induce adaptive immune responses [72]. IC50 threshold 500 nM or lower values were suggested to selects strong binding affinity between MHC class protein and epitopes [73]. When MHC molecules binds to epitopes, infected cell presents itself as an antigen-presenting cell for self-destruction. For better sensitivity, CD8+ T-cell epitopes were generated from both structural and non-structural proteins because both types of proteins will be processed by infected cells in the cytoplasm. Whereas structural proteins are of interest for CD4+ T-cell epitopes, as it might provide help to cognate interaction [74–76]. In this study, we selected CD4+ T-cell epitopes on the basis of CD8+ T-cell epitope core sequences to find out the best T-cell epitopes which might provide an immune response for both kinds of MHC classes. Twelve 15-mer MHC class II epitopes were selected which have core sequences of four 9-mer epitopes of MHC class I, and all four CD8+ cell epitopes belong to surface glycoprotein (Table 3). To understand demographic coverage of epitopes, population coverage analysis was performed through selected T-cell epitopes, and analysis revealed that approximately 89.44% and 93% average coverage can be achieved for world population and population of ethnic groups respectively (Supplementary File 2, Table S3). A particular antigen induces distinct class or subclass of antibodies such as schistosomiasis and filariasis induced a mixed response of IgE and IgG [77,78]. Predicted epitopes may or may not carry key feature of proteins because most of prediction methods ignored epitope and receptor interaction which might lead to produce an antibody of no use [60]. Therefore, sequence and structure-based dual approaches were used to identify distinct class of B-cell epitopes and 14 non-toxic, non-cross-reactive, antigenic, and immunogenic B-cell epitopes were identified of different length

(Table 4). The molecular docking approach was used to validate the interaction of three T-cell epitopes to most frequently occurring twenty two MHC allele's structures (Table 5). Similarly, the top five B-cell epitopes were used to explore peptide interaction with two different kind of B-cell receptor proteins (5DRW and 1K1F). First protein structure, 5DRW was considered to evaluate binding affinity of peptides to BCR antibody light chain, whereas 1K1F was a Bcr-Abl oncoprotein and formed a tetramer through oligomerization. Monomer of 1K1F protein provided a basis to design an inhibitors to disrupt Bcr-Abl oligomerization [61]. Therefore, 1K1F was considered as control to compare the peptide binding affinity to B-cell receptor Fab region binding affinity (Table 6). All peptides showed higher binding affinity to 5DRW than 1K1F except peptide KTVGELGDVRE (Supplementary File 1, Table S8). Docked complex of 1K1F and peptides was contained good interaction and total energy ( $-37.47$ , and  $-1490.12$ ) and six hydrogen bonds (THR2-PRO125, VAL3-SER127, GLU11-ASN144, GLY7-SER168, GLY7-SER182). Hydrogen bond visualization of 1K1F protein and peptide were given in Fig. 2.

In gene expression profiling of MHC genes and alleles, distinct gene expression patterns were observed for SARS-CoV-2 infection. Total, 92 gene variants of HLA-A\*24 were expressed in both cell lines and very low or no expression of HLA-A\*24 gene variants were found in A549 cell line. In literature, prevalence of HLA-A\*24 alleles was suggested as risk factors for severe H1N1 infection [79], and HLA-A\*24:02 alleles were also reported to increase diabetes-associated risk together with HLA-B\*39:01 gene [80]. A HLA-C allele, HLA-C\*03:03, associated with male infertility was also highly expressed in SARS-CoV-2 infected NHBE cell lines. The performed study on semen quality was reported that HLA-C\*03:03 allele expression was increased two fold in human papillomavirus virus infected individuals [81]. Three genetic variant of HLA-B\*08:01 genes, myasthenia gravis autoimmune disease characterized by muscle weakness and abnormal fatigability were also highly expressed in NHBE cell lines [82]. In present, a combination of malaria and AIDS drugs are in use for the treatment of COVID-19. So, it would be interesting to explore malaria and HIV associated MHC class allele's in SARS-CoV-2 infected samples. Therefore, all the MHC class genes were analysed and filtered, and a list of expressed MHC class alleles for malaria and HIV were generated by using IEDB resources along with expression counts for human cell lines (Supplementary File 1, Table S3). HLA-A\*02:01:131, HLA-A\*02:01:160 and HLA-A\*03:01:01:02 N were expressed for malaria and HIV in NHBE cell lines. SARS related two HLA genes (HLA-A\*23:01:03, HLA-A\*23:01:31) were expressed, and 34 HLA-A\*24:02 were expressed for HIV gag polyprotein in NHBE cell lines. In our analysis, HLA-A\*23:01 HLA-A\*24:02 and HLA-A\*02:01 were predicted for proposed T-cell epitopes.

## 5. Conclusion

This study has high scientific relevance to understand induced immune responses upon SARS-CoV-2 infection. Performed study has provided the extremely useful information about the expressed region of SARS-CoV-2 genome, expression profiling and differential gene expression of innate immune system genes and MHC class gene and alleles, T- and B- cell epitopes, and molecular interaction of identified epitopes to receptors. Expressed regions of SARS-CoV-2 genome and putative expressed targets of human immune response will facilitate vaccine related research studies. Proposed epitopes are possessing T- and B-cell selectivity, nontoxicity, higher population coverage, and significant interaction with MHC classes, and B-cell receptors. However, the presented list of T- and B-cell epitopes is an outcome of computational analysis. But, all the epitopes were identified from transcriptome data of SARS-CoV-2 infection in human cell-lines. Therefore, these epitopes have high potential to reflect SARS-CoV-2 immune response in human, and become vaccine candidates after experimental validation.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2020.09.019>.

## Data availability

All the used RNA-seq data is available at NCBI SRA under the project accession number PRJNA615032.

## Author's contributions

S.K., conceptualized the study and drafted the manuscript. S.K., V.B. and S.S. analyzed the data. S.C. and S.G. involved in data curation, review and editing of manuscripts. All authors read and approved the manuscript.

## Declaration of Competing Interest

Authors have declared that they have no competing interest.

## Acknowledgment

We are grateful to the Director NIAB for providing support to carry out the study. Bioinformatics facility, NIAB is acknowledged for the bioinformatics data analysis. We would also thankful to Aakash Chawade and Pallavi Chauhan for useful discussion and suggestions.

## References

- [1] Surveillance Case Definitions for Human Infection with Novel Coronavirus (nCoV): Interim Guidance v1, <https://apps.who.int/iris/handle/10665/330376>, (January 2020).
- [2] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet* 395 (10223) (2020) 507–513.
- [3] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, et al., Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin. B* 10 (5) (2020) 766–788.
- [4] D. Stadlbauer, F. Amanat, V. Chromikova, K. Jiang, S. Strohmeier, G.A. Arunkumar, J. Tan, D. Bhavsar, C. Capuano, E. Kirkpatrick, et al., SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup, *Curr. Protoc. Microbiol.* 57 (1) (2020) e100.
- [5] W. Liu, A. Fontanet, P.-H. Zhang, L. Zhan, Z.-T. Xin, L. Baril, F. Tang, H. Lv, W.-C. Cao, Two-year prospective study of the humoral immune response of patients with severe acute respiratory syndrome, *J. Infect. Dis.* 193 (6) (2006) 792–795.
- [6] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (7798) (2020) 270–273.
- [7] A.N. Zakhartchouk, Q. Liu, M. Petric, L.A. Babiuik, Augmentation of immune responses to SARS coronavirus by a combination of DNA and whole killed virus vaccines, *Vaccine* 23 (35) (2005) 4385–4391.
- [8] A.E. Thompson, The immune system, *JAMA* 313 (16) (2015) 1686.
- [9] A. Sette, J. Fikes, Epitope-based vaccines: an update on epitope identification, vaccine design and delivery, *Curr. Opin. Immunol.* 15 (4) (2003) 461–470.
- [10] S. Gras, S.R. Burrows, S.J. Turner, A.K. Sewell, J. McCluskey, J. Rossjohn, A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned, *Immunol. Rev.* 250 (1) (2012) 61–81.
- [11] M.V. Larsen, A. Lelic, R. Parsons, M. Nielsen, I. Hoof, K. Lamberth, M.B. Loeb, S. Buus, J. Bramson, O. Lund, Identification of CD8+ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL, *PLoS One* 5 (9) (2010) e12697, <https://doi.org/10.1371/journal.pone.0012697>.
- [12] S. Chakraborty, R. Chakravorty, M. Ahmed, A. Rahman, T.M. Waise, F. Hassan, M. Rahman, S. Shamsuzzaman, A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions, *In Silico Biol.* 10 (5–6) (2010) 235–246.
- [13] M.A. Hasan, M.A. Khan, A. Datta, M.H. Mazumder, M.U. Hossain, A comprehensive immunoinformatics and target site study revealed the corner-stone toward chikungunya virus treatment, *Mol. Immunol.* 65 (1) (2015) 189–204.
- [14] A.R. Oany, T. Pervin, M. Mia, M. Hossain, M. Shahnaj, S. Mahmud, K.M.K. Kibria, Vaccinomics approach for designing potential peptide vaccine by targeting *Shigella* spp. Serine protease autotransporter subfamily protein SigA, *J Immunol Res* 2017 (2017) 6412353.
- [15] D. Blanco-Melo, B.E. Nilsson-Payant, W.-C. Liu, R. Møller, M. Panis, D. Sachs, R.A. Albrecht, tenOever BR: SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems, *bioRxiv* (2020), <https://doi.org/10.1101/2020.03.24.004655>.
- [16] [https://ncbi.github.io/sra-tools/install\\_config.html](https://ncbi.github.io/sra-tools/install_config.html).
- [17] <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- [18] <https://multiqc.info/>.
- [19] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequencing data, *Bioinformatics* 30 (15) (2014) 2114–2120.
- [20] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods* 12 (4) (2015) 357–360.
- [21] <http://www.htslib.org/>.
- [22] <https://bedtools.readthedocs.io/en/latest/>.
- [23] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, et al., De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis, *Nat. Protoc.* 8 (8) (2013) 1494–1512.
- [24] <https://github.com/TransDecoder/TransDecoder/wiki>.
- [25] N.L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification, *Nat. Biotechnol.* 34 (5) (2016) 525–527.
- [26] H. Pimentel, N.L. Bray, S. Puente, P. Melsted, L. Pachter, Differential analysis of RNA-seq incorporating quantification uncertainty, *Nat. Methods* 14 (7) (2017) 687–690.
- [27] K. Breuer, A.K. Foroushani, M.R. Laird, C. Chen, A. Sribnaia, R. Lo, G.L. Winsor, R.E.W. Hancock, F.S.L. Brinkman, D.J. Lynn, InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation, *Nucleic Acids Res.* 41 (Database issue) (2013) D1228–D1233.
- [28] M.V. Larsen, A. Lelic, R. Parsons, M. Nielsen, I. Hoof, K. Lamberth, M.B. Loeb, S. Buus, J. Bramson, O. Lund, Identification of CD8+ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL, *PLoS One* 5 (9) (2010) e12697.
- [29] D.R. Flower, I.K. Macdonald, K. Ramakrishnan, M.N. Davies, Doytchinova IA: computer aided selection of candidate vaccine antigens, *Immunome Res.* 6 (Suppl. 2) (2010) S1, <https://doi.org/10.1186/1745-7580-6-S2-S1>.
- [30] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D.K. Wheeler, A. Sette, B. Peters, The immune epitope database (IEDB): 2018 update, *Nucleic Acids Res.* 47 (D1) (2019) D339–D343.
- [31] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G.P. Raghava, In silico approach for predicting toxicity of peptides and proteins, *PLoS One* 8 (9) (2013) e73957.
- [32] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (W1) (2017) W24–W29.
- [33] J. Ponomarenko, H.H. Bui, W. Li, N. Fusseder, P.E. Bourne, A. Sette, B. Peters, ElliPro: a new structure-based tool for the prediction of antibody epitopes, *BMC Bioinform.* 9 (2008) 514.
- [34] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (16) (2009) 2785–2791.
- [35] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31 (2) (2010) 455–461.
- [36] M. Kurcinski, M. Jamroz, M. Blaszczyk, A. Kolinski, S. Kmiecik, CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site, *Nucleic Acids Res.* 43 (W1) (2015) W419–W424.
- [37] T.D. Goddard, C.C. Huang, T.E. Ferrin, Software extensions to UCSF chimera for interactive visualization of large molecular assemblies, *Structure* 13 (3) (2005) 473–482.
- [38] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng.* 8 (2) (1995) 127–134.
- [39] Z. Chen, X. Zhang, C. Peng, J. Wang, Z. Xu, K. Chen, J. Shi, W. Zhu, D3Pockets: a method and web server for systematic analysis of protein pocket dynamics, *J. Chem. Inf. Model.* 59 (8) (2019) 3353–3358.
- [40] J. Trowsdale, J.C. Knight, Major histocompatibility complex genomics and human disease, *Annu. Rev. Genomics Hum. Genet.* 14 (1) (2013) 301–323.
- [41] A. Oeckinghaus, S. Ghosh, The NF- $\kappa$ B family of transcription factors and its regulation, *Cold Spring Harb. Perspect. Biol.* 1 (4) (2009) a000034.
- [42] S.C. Sun, J.H. Chang, J. Jin, Regulation of nuclear factor- $\kappa$ B in autoimmunity, *Trends Immunol.* 34 (6) (2013) 282–289.
- [43] S.C. Sun, Non-canonical NF- $\kappa$ B signaling pathway, *Cell Res.* 21 (1) (2011) 71–85.
- [44] D.P. Brazil, B.A. Hemmings, Ten years of protein kinase B signalling: a hard Akt to follow, *Trends Biochem. Sci.* 26 (11) (2001) 657–664.
- [45] H. Guo, P. German, S. Bai, S. Barnes, W. Guo, X. Qi, H. Lou, J. Liang, E. Jonasch, G.B. Mills, Z. Ding, The PI3K/AKT pathway and renal cell carcinoma, *J. Genet. Genomics* 42 (7) (2015) 343–353.
- [46] B.D. Manning, L.C. Cantley, AKT/PKB signaling: navigating downstream, *Cell* 129 (7) (2007) 1261–1274.
- [47] C. Widmann, S. Gibson, M.B. Jarpe, G.L. Johnson, Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human, *Physiol. Rev.* 79 (1) (1999) 143–180.
- [48] K.L. Pierce, L.M. Luttrell, R.J. Lefkowitz, New mechanisms in heptahelical receptor signaling to mitogen activated protein kinase cascades, *Oncogene* 20 (13) (2001) 1532–1539.
- [49] R. Seger, E.G. Krebs, The MAPK signaling cascade, *FASEB J.* 9 (9) (1995) 726–735.
- [50] D. Willoughby, D.M. Cooper, Organization and Ca<sup>2+</sup> regulation of adenylyl cyclases in cAMP microdomains, *Physiol. Rev.* 87 (3) (2007) 965–1010.
- [51] X. Cheng, Z. Ji, T. Tsalkova, F. Mei, Epac and PKA: a tale of two intracellular cAMP receptors, *Acta Biochim. Biophys. Sin. Shanghai* 40 (7) (2008) 651–662.
- [52] A.M. Hofer, K. Lefkimiatis, Extracellular calcium and cAMP: second messengers as “third messengers”? *Physiology (Bethesda)* 22 (2007) 320–327.
- [53] A.C. Ferguson-Smith, Y.F. Chen, M.S. Newman, L.T. May, P.B. Sehgal, F.H. Ruddle, Regional localization of the interferon-beta 2/B-cell stimulatory factor 2/hepatocyte stimulating factor gene to human chromosome 7p15–p21, *Genomics* 2 (3)

- (1988) 203–208.
- [54] S. Cohen, W.J. Doyle, D.P. Skoner, Psychological stress, cytokine production, and severity of upper respiratory illness, *Psychosom. Med.* 61 (2) (1999) 175–180.
- [55] M. Ghosh, V. Aguirre, K. Wai, H. Felfly, W.D. Dietrich, D.D. Pearse, The interplay between cyclic AMP, MAPK, and NF- $\kappa$ B pathways in response to proinflammatory signals in microglia, *Biomed. Res. Int.* 2015 (2015) 308461.
- [56] M.C. Mendoza, E.E. Er, J. Blenis, The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation, *Trends Biochem. Sci.* 36 (6) (2011) 320–328.
- [57] H.-H. Bui, J. Sidney, K. Dinh, S. Southwood, M.J. Newman, A. Sette, Predicting population coverage of T-cell epitope-based diagnostics and vaccines, *BMC Bioinform.* 7 (2006) 153.
- [58] [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200502-covid-19-sitrep-103.pdf?sfvrsn=d95e76d8\\_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200502-covid-19-sitrep-103.pdf?sfvrsn=d95e76d8_4).
- [59] S.J. Darnell, M. Riese, E.G. Edlunde, F.R. Blattner, Enhancing B-cell epitope predictions by integrating protein sequence and structural bioinformatics, *Biophys. J.* 106 (2) (2014) 207a.
- [60] L. Potocnakova, M. Bhide, L.B. Pulzova, An introduction to B-cell epitope mapping and in silico epitope prediction, *J Immunol Res* 2016 (2016) 6760830, <https://doi.org/10.1155/2016/6760830>.
- [61] X. Zhao, S. Ghaffari, H. Lodish, V.N. Malashkevich, P.S. Kim, Structure of the Bcr-Abl oncoprotein oligomerization domain, *Nat. Struct. Biol.* 9 (2) (2002) 117–120.
- [62] N.L. Tilston-Lunel, G.O. Acrani, R.E. Randall, R.M. Elliott, Generation of recombinant Oropouche viruses lacking the nonstructural protein NSm or NSs, *J. Virol.* 90 (5) (2016) 2616.
- [63] T. Tanaka, W. Kamitani, M.L. DeDiego, L. Enjuanes, Y. Matsuura, Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA, *J. Virol.* 86 (20) (2012) 11128–11137.
- [64] I.G. Goodfellow, L.O. Roberts, Eukaryotic initiation factor 4E, *Int. J. Biochem. Cell Biol.* 40 (12) (2008) 2675–2680.
- [65] R.L. Graham, A.C. Sims, S.M. Brockway, R.S. Baric, M.R. Denison, The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication, *J. Virol.* 79 (21) (2005) 13399.
- [66] M. Kikkert, Innate immune evasion by human respiratory RNA viruses, *J. Innate Immun.* 12 (1) (2020) 4–20.
- [67] Y. Gao, N. Goonawardane, J. Ward, A. Tuplin, M. Harris, Multiple roles of the non-structural protein 3 (nsP3) alphavirus unique domain (AUD) during chikungunya virus genome replication and transcription, *PLoS Pathog.* 15 (1) (2019) e1007239, <https://doi.org/10.1371/journal.ppat.1007239>.
- [68] L.G. Gardinassi, C.O.S. Souza, H. Sales-Campos, S.G. Fonseca, Immune and metabolic signatures of COVID-19 revealed by transcriptomics data reuse, *Front. Immunol.* 11 (1636) (2020).
- [69] D. Blanco-Melo, B.E. Nilsson-Payant, W.-C. Liu, S. Uhl, D. Hoagland, R. Möller, T.X. Jordan, K. Oishi, M. Panis, D. Sachs, et al., Imbalanced host response to SARS-CoV-2 drives development of COVID-19, *Cell* 181 (5) (2020) 1036–1045 e1039.
- [70] E.M. Cottam, M.C. Whelband, T. Wileman, Coronavirus NSP6 restricts autophagosome expansion, *Autophagy* 10 (8) (2014) 1426–1441.
- [71] A.J.W. te Velthuis, S.H.E. van den Worm, E.J. Snijder, The SARS-coronavirus nsp7 + nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension, *Nucleic Acids Res.* 40 (4) (2012) 1737–1747.
- [72] A.N. Ilinskaya, M.A. Dobrovol'skaia, Understanding the immunogenicity and antigenicity of nanomaterials: past, present and future, *Toxicol. Appl. Pharmacol.* 299 (2016) 70–77.
- [73] W. Zhao, X. Sher, Systematically benchmarking peptide-MHC binding predictors: from synthetic to naturally processed epitopes, *PLoS Comput. Biol.* 14 (11) (2018) e1006457, <https://doi.org/10.1371/journal.pcbi.1006457>.
- [74] Y. Tian, A. Grifoni, A. Sette, D. Weiskopf, Human T cell response to dengue virus infection, *Front. Immunol.* 10 (2019) 2125, <https://doi.org/10.3389/fimmu.2019.02125>.
- [75] A. Sette, M. Moutaftsi, J. Moyron-Quiroz, M.M. McCausland, D.H. Davies, R.J. Johnston, B. Peters, M. Raffi-El-Idrissi Benhnia, J. Hoffmann, H.-P. Su, et al., Selective CD4+ T cell help for antibody responses to a large viral pathogen: deterministic linkage of specificities, *Immunity* 28 (6) (2008) 847–858.
- [76] S. Parvizpour, M.M. Pourseif, J. Razmara, M.A. Rafi, Y. Omid, Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches, *Drug Discov. Today* 25 (6) (2020) 1034–1042, <https://doi.org/10.1016/j.drudis.2020.03.006>.
- [77] P. Hagan, U.J. Blumenthal, D. Dunn, A.J.G. Simpson, H.A. Wilkins, Human IgE, IgG4 and resistance to reinfection with *Schistosoma haematobium*, *Nature* 349 (6306) (1991) 243–245.
- [78] E.A. Ottesen, F. Skvaril, S.P. Tripathy, R.W. Poindexter, R. Hussain, Prominence of IgG4 in the IgG antibody response to human filariasis, *J. Immunol.* 134 (4) (1985) 2707.
- [79] T. Hertz, C.M. Oshansky, P.L. Roddam, J.P. DeVincenzo, M.A. Caniza, N. Jovic, S. Mallal, E. Phillips, I. James, M.E. Halloran, et al., HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection, *Proc. Natl. Acad. Sci.* 110 (33) (2013) 13492–13497.
- [80] M.-L. Mikk, T. Heikkinen, M.I. El-Amir, M. Kiviniemi, A.-P. Laine, T. Härkönen, R. Veijola, J. Toppari, M. Knip, J. Ilonen, et al., The association of the HLA-A\*24:02, B\*39:01 and B\*39:06 alleles with type 1 diabetes is restricted to specific HLA-DR/DQ haplotypes in Finns, *HLA* 89 (4) (2017) 215–224.
- [81] P.I. Marques, J.C. Gonçalves, C. Monteiro, B. Cavadas, L. Nagirnaja, N. Barros, A. Barros, F. Carvalho, A.M. Lopes, S. Seixas, Semen quality is affected by HLA class I alleles together with sexually transmitted diseases, *Andrology* 7 (6) (2019) 867–877.
- [82] J. Varade, N. Wang, C.K. Lim, T. Zhang, Y. Zhang, X. Liu, F. Piehl, R. Matell, H. Cao, X. Xu, et al., Novel genetic loci associated HLA-B\*08:01 positive myasthenia gravis, *J. Autoimmun.* 88 (2018) 43–49.